

**APPLICATION FOR UNITED STATES LETTERS PATENT**

**INVENTORS:** Edward F. PATZ, Jr., Michael J. CAMPA, and Michael C. FITZGERALD

**TITLE:** SYSTEM AND METHOD FOR DIFFERENTIAL PROTEIN EXPRESSION AND A DIAGNOSTIC BIOMARKER DISCOVERY SYSTEM AND METHOD USING SAME

**ATTORNEYS:** FLESHNER & KIM, LLP  
&  
**ADDRESS:** P. O. Box 221200  
Chantilly, VA 20153-1200

**DOCKET NO.:** DUKE - 227

## BACKGROUND OF THE INVENTION

### 1. Field of the Invention

[1] The present invention relates generally to a system and method for determining differential protein expression, and a diagnostic biomarker discovery system and method that utilizes the same. In particular, the present invention relates to a system and method of obtaining and analyzing cell or specimen protein profiles so as to correlate protein patterns with clinical parameters and manifestations of disease in the discovery of specific biomarkers

### 2. Background of the Related Art

[2] There is a continuing need for innovative strategies that allow early detection, diagnosis, treatment, monitoring and prognosis of diseases, such as cancer and other biological conditions, and inability to tolerate certain medications or treatments. While current non-invasive radiologic studies and laboratory tests play an integral role in the evaluation of diseases and biological conditions, there are clear limitations for early detection and specific diagnosis. For example, early detection efforts and screening trials for various cancers, even targeted at high risk individuals, have often been ineffectual. See, for example: Fontana, R.S. et al., "Early Lung Cancer Detection: Results of the Initial (Prevalence) Radiologic and Cytologic Screening in the Mayo Clinic Study", *Am. Rev. Respir. Dis.* 130: 561-565 (1984); Berlin, N.I., et al., "The National Cancer Institute Cooperative Early Lung Cancer Detection Program: Results of the Initial Screen (Prevalence)", *Am. Rev. Respir. Dis.* 130: 545-549 (1984); Kubik, A. and Polak, J., "Lung Cancer Detection: Results of a Randomized Prospective Study in Czechoslovakia", *Cancer* 57: 2427-2437 (1986); Fontana, R.S. et al., "The Mayo Lung Project for Early Detection and Localization of Bronchogenic Carcinoma: A Status Report", *Chest* 67: 511-522 (1975); Tockman, M.S., "Survival and Mortality from Lung Cancer in a Screened Population. The Johns Hopkins Study", *Chest* 89 (suppl.): 324S-325S (1986); Fontana, R.S. et al., "Screening for Lung Cancer. A Critique of the Mayo Lung Project", *Cancer* 67: 1,155-1,164 (1991); and Marcus, P.M. et al., "Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up", *J.*

*Natl. Cancer Inst.* 92: 1,308-1,315 (2000). Thus an alternative approach to early detection, accurate diagnosis and characterization of disease, and prognosis is needed.

[3] In recent years, it has been demonstrated that certain substances, including proteins, referred to as biomarkers, are expressed differentially in the diseased tissue and specimens versus the normal tissue and specimens. For example, it is believed that a differentially expressed protein that is found to be present in diseased tissue of many patients, while being absent in the normal tissue, is a candidate biomarker for that disease. Rasmussen et al., *Electrophoresis* 15:406-416 (1994); Hong Ji et al., *Electrophoresis* 15:391-405 (1994); Prasad S.C. et al., *Int. J. Oncology* 14:529-534 (1999); Soldes O.S. et al., *British J. of Cancer* 79(3/4):595-603 (1999). Biomarkers, hence, provide an additional measure for medical diagnosis and prognosis.

[4] Often, however, a single biomarker may be insufficient for accurate diagnosis of disease onset, and the search continues for the optimal panel of biomarkers that together can provide a profile for a given disease or condition at various stages of its pathology. Emmert-Buck, M.R. et al., *Mol. Carcinogenesis* 27:158-165 (2000). It is envisioned that a combination of biomarker information, as well as the traditional indicia of medical diagnoses, can provide a more accurate and early detection system.

[5] In some instances, the diagnostic and prognostic problems associated with various diseases and conditions are made more complicated by the fact that not enough biomarkers for these diseases have been found yet. Hence, there is a need in the art to rapidly identify such biomarkers. But even when a panel of biomarkers are known for a given disease or condition, no integrated system is yet available that accurately and expediently detects and analyzes the protein profile of a given patient so that a timely diagnosis, preferably at the onset of the disease or condition, can be made and the needed course of treatment started at an early stage when the disease or condition is more likely to be responsive to treatment.

[6] The above references are incorporated by reference herein where appropriate for appropriate teachings of additional or alternative details, features and/or technical background.

### **SUMMARY OF THE INVENTION**

[7] In view of the above described problems and limitations of the prior art, it is an object of the invention to solve at least the above problems and limitations by providing at least the advantages described hereinafter.

[8] The present invention relates to a database of protein patterns associated with diseases or other biological conditions.

[9] The present invention also relates to a database that stratifies patients having common diagnosis and clinical outcomes.

[10] The present invention also relates to a database that contains patient clinical information, images, mass spectrometer spectra and data analysis.

[11] The present invention also relates to an algorithm for analyzing protein expression data.

[12] The present invention also relates to an artificial neural network for analyzing protein expression data.

[13] The present invention also relates to an algorithm for recognizing informative patterns of protein expression that can be correlated with clinical parameters and manifestations of disease.

[14] The present invention also relates to a system and methodology for creating a comprehensive protein profile.

[15] The present invention also relates to a system and methodology for identifying protein patterns associated with predetermined biological characteristics.

[16] The present invention also relates to a system and methodology for identifying protein patterns associated with predetermined clinical parameters.

[17] The present invention also relates to a system and methodology for identifying protein patterns associated with predetermined medical conditions.

[18] The present invention also relates to a system and methodology for identifying protein patterns associated with predetermined diseases.

[19] The present invention also relates to a system and methodology for predicting the existence or non-existence of at least one predetermined biological characteristic.

[20] The present invention also relates to a system and methodology for predicting the presence of disease in an animal body, such as a mammal.

[21] The present invention also relates to a system and methodology for rapidly identifying proteins associated with disease or other biological conditions that can be used as biomarkers in diagnostic applications.

[22] The present invention also relates to a system and methodology for using a biomarker protein as a non-invasive imaging target for one or more sites of diseased cells in a mammalian body.

[23] The present invention also relates to a system and methodology for using biomarker proteins as a therapeutic target for treatment of disease or other biological conditions.

[24] The present invention also relates to a system and methodology for discovering proteins that are useful as imaging or therapeutic targets of disease.

[25] The present invention also relates to protein biomarkers for monitoring the course of a disease, and for determining appropriate therapeutic intervention.

[26] The present invention also relates to a system and methodology for using biomarker proteins as targets for drug delivery systems in a mammalian body in order to enhance drug efficacy.

[27] Additional advantages, objects, and features of the invention will be set forth in part in the description which follows, and in part will become apparent to those having ordinary skill in the art upon examination of the following or may be learned from practice of the invention. The objects and advantages of the invention may be realized and attained as particularly pointed out in the appended claims.

## BRIEF DESCRIPTION OF THE DRAWINGS

[28] The invention will be described in detail with reference to the following drawings,

wherein:

[29] Figure 1 is a block diagram of a cell protein profiling and diagnostic system, in accordance with the present invention;

[30] Figure 2A is a flowchart of one preferred method of identifying and storing cell protein patterns using the system of Fig. 1, in accordance with the present invention;

[31] Figure 2B is a flowchart of one preferred diagnosing method using the system of Fig. 1, in accordance with the present invention;

[32] Figure 2C is a flowchart of one preferred method of preparing a tissue sample for protein fractionation, in accordance with the present invention;

[33] Figure 3 is a graph showing representative spectra of tumor and normal lung lysates analyzed on a cation exchange surface, in accordance with the present invention;

[34] Figure 4 is a graph showing representative spectra of tumor and normal lung lysates analyzed on an anion exchange surface; and

[35] Figure 5 is a graph showing representative spectra of tumor and normal lung lysates analyzed on an immobilized metal infinity surface.

## DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[36] The present invention provides an apparatus and methodology for rapidly identifying new biomarkers, generating a comprehensive database of biomarkers and other indicia for medical diagnosis and prognosis, generating substantially complete protein profiles for a given population, and allowing generation and comparison of the protein profile of a given individual against the population profile, thereby detecting the differences that point to the presence or absence of disease or other biological conditions.

[37] In a preferred embodiment of the invention, a tissue sample or specimen, such as urine, blood, or other readily obtainable and minimally invasive biological sample, is obtained

from the patient. The sample is used to generate cell or specimen lysates. Any methodology, including the ones described herein below, may be used to make cell or specimen lysates.

[38] Next, the total complex protein composition is fractionated into sub-groups. Any methodology may be used to fractionate the proteins into sub-groups, as long as the complexity of the original protein mixture is reduced. Protein fractionation may be done based on any given property, e.g. size, charge, isoelectric point, or hydrophobicity, as long as the fractions obtained are sufficiently reduced in complexity to permit detection by mass spectrometry of the greatest possible proportion of all the proteins in the fraction.

[39] It is advisable to use one or several different types of separation steps in order to fractionate the cell lysates prior to mass spectrometric analysis. Such chromatographic steps include, but are not limited to, the following: normal and reversed-phase high performance liquid chromatography (HPLC), ion-exchange chromatography, size exclusion chromatography, 1D or 2D gel electrophoresis, isoelectric focusing, and capillary electrophoresis. Experimental results have shown that the use of reversed-phase HPLC to fractionate cell lysates can affect the number and distribution of proteins detected by spectrometry. When the eluant from the reversed-phase HPLC separation is subjected to spectrometry (e.g. MALDI) analysis, an increased number of proteins are clearly detected.

[40] The number of fractions generated for analysis may vary based on the given particulars at hand, described below. It is expected, however that the fractions generated would contain as few as less than 10 to as high as 1,500 proteins. In general, HPLC will generate more complex fractions than a gel fractionation method, such as 2D gel electrophoresis. However, since the proportion of fractioned proteins that are analyzable by mass spectrometry will differ depending on the fractionation method used, the most effective method will involve more than one fractionation scheme.

[41] After fractionating the total cell or specimen protein content into sub-groups or fractions, each protein fraction or sub-group is then analyzed by mass spectrometry using, for example, Matrix Assigted Laser Desorption/Ionization (MALDI) or Surface-Enhanced Laser Desorption Ionization (SELDI) time-of-flight mass spectrometry. Without fractionation, mass

spectrometry analysis of complex protein mixtures such as those in whole cell lysates can be compromised due to the fact that different peptide and protein analytes can experience preferential desorption/ionization in the mass spectrometry process. In some cases, signal suppression effect can be so severe that certain peptides and proteins are not detected in the presence of others.

[42] In designing the present invention, the initial mass spectrometry experiments of tumor cell lysates were carried out using mass spectrometry samples directly from the cell or specimen lysates without any fractionation step (*see* Example 1 below). This, however, typically allowed detection on the order of 30-50 peptides and proteins, an estimated less than 1% of the total protein content of the cell. To visualize many more proteins and produce the most comprehensive disease profile possible, the protein fractionation step was devised to be carried out prior to mass spectrometry analysis, so that each fraction will generate a diverse protein spectrum. The fractionation step, which makes use of a variety of separation techniques, increases the number of proteins identified in the complete expression profile of the lysate.

[43] The data output from the mass spectrometry is an array, or spectrum, of peaks with each peak representing a protein or group of proteins present in a given sample. The location of any given peak on the x-axis is related to the molecular mass and charge of the protein, while the height of the peak is associated with the relative abundance of the protein ion. For a given set of experimental conditions, the spectrum represents a molecular profile of the protein sub-group or fraction of the expressed proteins in a given specimen.

[44] By comparing the protein spectra between different specimens or between the specimen and the established control(s), differences between them can be ascertained. For example, by comparing the spectrum of healthy tissue to a spectrum of diseased tissue from the same patient, differences in the expression of specific proteins can be detected. Hence, a differentially expressed protein or proteins that are found in diseased tissue of many patients, while being absent in the normal tissue, is a candidate biomarker for that disease. Similarly, the differences between the protein profile of a given patient and the profile generated from studying a population to which the patient is related, are indicative of the presence or absence

of a biomarker, which can assist in the diagnosis and/or prognosis of a disease or biological condition.

[45] The present invention makes use of neural networks and other analysis techniques to determine which proteins are common to patients with the same disease. In addition, the data is mined to determine the differences in protein expression between the diseased/abnormal and normal subjects (and other diseases or abnormalities), and thus create a series of patterns of protein expression unique to that specific disease or biological condition. Individual proteins found in specific diseases or abnormalities, and not found in normal specimens, can be identified as possible therapeutic targets.

[46] This creation of protein patterns for specific diseases or other biological conditions will allow the system described herein to analyze any unknown specimen and determine the diagnosis with prognostic and therapeutic implications.

[47] Figure 1 is a block diagram of a cell or specimen protein profiling and diagnostic system 100, in accordance with the present invention. The system comprises a protein fractionation unit 110, a mass spectrometer 120, a cell protein data processing unit 130, an input unit 140 and a protein profile database 150.

[48] The system 100 is used to create substantially complete protein profiles for samples, identify protein patterns in the cell protein profiles that are associated with subject characteristics, such as biological conditions and diseases, and storing these protein profiles and identified protein patterns for later use in diagnostic applications.

[49] The operation of the system 100 will be further described in connection with Figs. 2B and 2C, which are flowcharts of a preferred method of identifying and storing disease protein patterns, and a preferred diagnosing method, respectively. The method of Fig. 2B begins at step 200, where a tissue sample is obtained from a subject. The type of tissue sample selected depends on the type of disease protein pattern that one wants to identify. However, the tissue sample is typically not composed of a homogeneous population of one cell type. For example, a specimen of lung tumor is composed of cancer cells, normal lung cells, blood cells, endothelial cells, etc. However, tumor specimens from two different subjects may contain similar

populations of cells. This could be ascertained by the examination of stained thin sections of the tissue sample being analyzed.

[50] At step 210, the protein fractionation unit 110 fractionates proteins from the tissue sample into protein subgroups. A tissue sample can contain tens of thousands of different proteins, and possibly over one hundred thousand distinct proteins if post-translational modification is performed. Mass spectrometers currently available do not have the resolution required to visualize every distinct protein in a tissue sample.

[51] Accordingly, one aspect of the present invention is the recognition that fractionating the proteins found in the tissue sample into multiple subgroups, and performing mass spectrometry on each protein subgroup, will increase the number of proteins detected in a given sample.

[52] Any technique can be used by the protein fractionation unit 100 to fractionate the proteins found in the tissue sample into protein subgroups. For example, the fractionation can be done by size, charge, isoelectric point or hydrophobicity. Whatever technique is used, the fractions obtained must be sufficiently reduced in complexity to permit detection, by mass spectrometry, of the largest possible proportion of all the proteins contained in the fraction.

[53] A preferred method for performing the protein fractionation is analytical reversed-phase high performance liquid chromatography (RP-HPLC). One example of an instrument that can be used to perform the analytical RP-HPLC is a Dynamax SD-200 solvent delivery system, and a Dynamax Variable Wavelength UV/Visible Absorbance Detector.

[54] Analytical RP-HPLC is preferably performed on a C4 Vydac column (0.46x 15.0cm, 300angstroms) at a flow rate of 1mL-min. Separations are preferably performed using linear gradients of Buffer B in A (Buffer A=0.1% TFA in water, and Buffer B=90% acetonitrile in water containing 0.09% TFA). A 0 to 67% gradient of Buffer B in A is preferably used for the separation. However, other gradient schemes and buffer compositions can also be used.

[55] A fractionation scheme such as analytical RP-HPLC will generate 20 fractions. Thus, assuming 37,000 different proteins are present in the tissue sample, each fraction will have approximately 1,850 proteins.

[56] A gel-base fractionation technique is able to generate more fractions than the analytical RP-HPLC technique. For a 1D gel that is 10cm long, one can obtain from 100-1,000 fractions, depending on whether the fraction is 1mm or 0.1mm in length. The number of fractions increases dramatically with a 2D gel to 10,000-100,000 fractions, depending on the size of the spot analyzed (1.0 or 0.1mm on a side). Although not all spots will contain protein, one still obtains a large number of fractions.

[57] As discussed above, fractionation will typically be able to generate fractions that contain as few as less than 10 proteins per fraction, to as many as over 1,500 proteins per fraction. In general, analytical RP-HPLC will generate more complex fractions than gel fractionation. However, since the proportion of a fractionated proteins that are analyzable by mass spectrometry will differ depending on the fractionation method used, the most affective protein fractionation method may involve using more than one fractionation technique. Other fractionation techniques that can be used include, but are not limited to, normal HPLC ion-exchange chromatography, size exclusion chromatography, and capillary electrophoresis.

[58] Clearly, to avoid protein degradation, appropriate steps should be taken to preserve the protein content of the samples. The tissue sample should be prepared as soon as possible after it is obtained, or stored in liquid nitrogen or otherwise at approximately -80°C. Once the proteins and the tissue sample are fractionated, the protein fractions should be analyzed, or stored in liquid nitrogen or otherwise at approximately -80°C.

[59] At step 220, mass spectrometry is performed on each protein subgroup that comes out of the fractionation process. The mass spectrometry is preferably performed using Matrix Assisted Laser Desorption/Ionization Time-Of-Flight (MALDI-TOF) mass spectrometry. However, a variety of other mass spectrometric methods such as SELDI and Electrospray Ionization (ESI) may also be used.

[60] Each protein sub-group is preferably prepared for MALDI-TOF mass spectrometry by combining approximately 1 $\mu$ L of the protein sub-group with approximately 30 $\mu$ L of MALDI substrate solution (or with solution appropriate for whatever mass

spectrometric procedure is used), which contains a saturated aqueous solution of sinapinic acid containing 50% acetonitrile and 0.1% trifluoracetic acid (TFA), or other matrices.

[61] The saturated solution of sinapinic acid is preferably prepared by adding solid sinapinic acid to a 50:50 (v/v) solution of water and acetonitrile with 0.1% (v/v) of TFA. The approximate ratio of (30:1) of MALDI substrate solution to protein lysate extract can be varied beyond this ratio on a case-by-case basis to effect an optimal concentration for MALDI-TOF mass spectrometry for a given situation.

[62] For each protein sub-group that is run through the mass spectrometer 120, a mass/amplitude spectrum is generated. Specifically, the time-of-flight data for a given protein in a mixture is translated into the mass/charge ratio for the protein, or  $m/z$ . Because the charge is typically assumed to be +1, the  $m/z$  values in a spectrum are considered to be equivalent to the molecular mass of the protein plus the mass of a proton (i.e., 1). The resulting data is in the form of a X-Y plot where peaks, representing individual proteins or groups of proteins, are arrayed along the x-axis at their respective  $m/z$  values. The height of each peak is proportional to the detector response and, hence, can be interpreted as the relative abundance of the protein ions contributing to the peak.

[63] At steps 230 and 240, the cell protein data processing unit 120 analyzes the mass spectra for each of the protein sub-groups to create a cell protein profile, and identifies protein patterns associated with subject characteristics. Subject characteristics typically include patient clinical information such as age, sex, disease, outcome, stage at presentation and response to therapy.

[64] The subject characteristics are input to the cell protein data processing unit 130 with input unit 140. Input unit 140 is suitably a computer that stores subject information.

[65] The cell protein data processing unit 130 obtains information regarding protein expression patterns that are specific to diseases by comparing the mass spectrometer spectra between specimens representing diseased and healthy states. The cell protein profiles and protein patterns identified by the cell protein data processing unit 130 are stored, at step 250, in the protein profile database 150. The database 150 preferably incorporates fields for entry of

spectra and for seamless integration of data analysis. Each database entry preferably contains patient clinical information, images (CT, PET radiographs), mass spectrometer spectra, and data analysis.

[66] Fig. 2B is a flowchart of one preferred diagnosing method, utilizing the system 100 of Fig. 1. Steps 300-330 are similar to steps 200-230 in the method of Fig. 2A, and thus will not be explained again.

[67] At step 340, the cell protein data processing unit compares the cell protein profile with the protein patterns previously identified and stored in the database 150. At step 350, the existence or non-existence of subject characteristics, such as biological conditions or diseases, are predicted by the cell protein data processing unit 130.

[68] The raw time-of-flight versus amplitude data received by the cell protein data processing unit 130 may consist of tens of thousands of individual measurements for each tissue sample analyzed. While it may be possible to obtain useful information regarding protein expression differences among very small groups of tissue samples with the naked eye, a thorough comparison among many hundreds of tissue samples is preferably performed with a computer algorithm that is executed by the cell protein profiling unit 130.

[69] Accordingly, the cell protein data processing unit 130 preferably utilizes an algorithm to identify the protein patterns associated with subject characteristics, such as predetermined medical conditions or diseases. The algorithm is preferably designed to recognize informative patterns of protein expression that may be correlated with clinical parameters and manifestations of disease. The algorithm is also preferably designed to identify proteins associated with disease that may be used as biomarkers in *in vitro* diagnostic applications, or as targets for non-invasive imaging or to guide the delivery of cytotoxic or therapeutic agents.

[70] The algorithm may be based on an Artificial Neural Network (ANN). Given  $N$  cases, the ANN is preferably trained on  $N-1$  cases, and then validated on the one case left out. This process is preferably repeated  $N$  times until each case has served as a validation case, and then all  $N$  results are combined. The resulting ANN analyzes each peak separately and attempts to predict if it originated from a diseased tissue sample or a normal tissue sample.

[71] When an ANN, as described above, was used on a data set with a total 248 peaks, a 93% sensitivity and a 61% specificity in identifying spectra as "disease" or "normal" was achieved. The sensitivity can be increased to approximately 95% by combining the original ANN with a second ANN based on a different molecular mass range. However, this additional classification step decreases the specificity to 58%.

[72] A second preferred algorithm uses all data points contained in a mass spectrometer spectrum, as opposed to using only the peaks identified by the mass spectrometer software. With this algorithm, the data are first filtered in order to produce a uniform base line amount among all sample spectra. Next, the sample data sets are put through a T-squared test to determine which bins are the most valuable in terms of their ability to separate the two sample sets (diseased and normal) of data.

[73] The test yields a P-value for each bin, which reflects the probability that the means of the two groups of data in that bin are equal. A very low P-value indicates that the two means are not close to each other, and thus that bin has a reasonable capability of separating the sample sets. The lower the P-value, the more separable the data is in that particular bin.

[74] Fig. 2C is a flowchart of a preferred method for preparing the tissue sample for protein fractionation, as part of steps 210 and 310 in the methods of Figs. 2A and 2B, respectively. The method begins at step 400, where the blood content of the tissue sample is reduced by incubating the tissue sample in 10mL PBS at approximately 4°C for approximately 30 minutes.

[75] Then, at step 410, a portion of the tissue sample is crushed in a protein extraction reagent. Specifically, a small portion of the cell sample (preferably 10-20mg wet weight) is preferably placed into a 1.5ml microcentrifuge tube containing 65µL Mammalian Protein Extraction Reagent (M-PER). The portion of the tissue sample is crushed in the M-PER preferably using a plastic microcentrifuge-sized pestle, and then shaken for approximately 10 minutes at approximately 40°C.

[76] Next, at step 420, insoluble material is removed by centrifugation at 16,000 x g at approximately 4°C for approximately 20 minutes. At step 430, the supernatant fraction is

stored, preferably in a clean microcentrifuge tube, in liquid nitrogen or otherwise at approximately -80°C until it is used.

## Examples

[77] The following examples are intended to further illustrate certain embodiments of the invention and are not intended to be limiting in nature.

### Example 1

[78] MALDI samples of tumor and normal cell lysates were prepared by combining 1  $\mu$ l of the unpurified cell lysate with 30  $\mu$ l of a saturated aqueous solution of sinapinic acid containing 50% acetonitrile and 0.1% trifluoracetic acid (TFA). Ultimately, 1-2  $\mu$ l of the resulting mixture was deposited on the MALDI sample stage, and the solvent was evaporated at room temperature. MALDI mass spectra were acquired on a Voyager DE Biospectrometry Workstation (PerSeptive Biosystems, Inc., Framingham, MA) in the linear mode using a nitrogen laser (337 nm).

[79] All mass spectra were collected in the positive-ion mode, and the spectra represent the sum of approximately 32 laser shots. The raw intensity versus time data was smoothed using a Savitsky-Golay smoothing routine prior to mass calibration using an internal standard. Using the simple MALDI sample preparation described above, approximately 30-50 peptides and proteins were detected, which is less than 1% of the total protein content of the cell. Interestingly, in this relatively small population of proteins, at least 1 protein was identified that appears unique to tumor cell lysates. These profiles can be used to accurately separate tumor from normal samples and other diseases based on their protein spectrum.

### Example 2

[79] One of the differences between SELDI and conventional MALDI-TOF is the ProteinChip™ technology for sample application. ProteinChips are available with a variety of

chemical surfaces, which permits the capture and analysis of whole classes of proteins based on their charge, hydrophobicity, or metal binding capability. The analysis of a biological specimen using just one surface may give information on 40-60 different proteins. By using a series of different surfaces and different wash conditions, it is possible to differentiate 500-1,000 proteins. However, sample preparation and analysis must be optimized for each ProteinChip surface and for each sample type.

[80] ProteinChip surfaces include cation exchange, anion exchange, reverse phase, and immobilized metal affinity capture. Protocols for binding sample to the surfaces and subsequent wash steps are developed much the same way as for column chromatography employing equivalent separation matrices. For example, initial studies using the cation exchange surface have been in a low pH buffer in order to maximize the number of proteins adsorbed to the surface. Potential disease-specific biomarkers identified in the screens can then be partially purified on the ProteinChip surface using wash buffers of progressively higher pH.

[81] Figure 3 shows representative spectra of tumor (top) and normal (bottom) lung lysates analyzed on a cation exchange surface (WCX-2). The numbers associated with the peaks are mass/charge ( $m/z$ ) values. Since the charge is +1, the values represent the molecular mass of each protein. The large peak at 22600 Da and the tumor lysate is absent in a normal lung tissue. Likewise, there are peaks at approximately 28,000 and 31,000 Da that present in the normal, but not the tumor. Following verification of these protein expression differences using several different tumor/normal tissue pairs, one can begin to isolate these proteins on the chip surface. Since the molecular masses determined by SELDI are very accurate, protein identity can often be achieved by simply searching web-based databases using the molecular mass value. If this is unsuccessful, the isolated protein can be digested with a protease and the resultant peptides separated on the SELDI and peptide fingerprint databases searched.

[82] In addition to protocols for the cation exchange surface, protocols for anion exchange (SAX-2) and immobilized metal infinity (IMAC-3) have been derived. Representative spectra from each are shown in Figs. 4 and 5, respectively.

[83] It is evident that each ProteinChip surface captures a different set of proteins, and each set displays tumor/normal protein expression differences. In order to survey the largest possible set of expressed proteins, all specimens are prefably analyzed using multiple ProteinChip surfaces.

[84] The foregoing embodiments and advantages are merely exemplary and are not to be construed as limiting the present invention. The present teaching can be readily applied to other types of apparatuses. The description of the present invention is intended to be illustrative, and not to limit the scope of the claims. Many alternatives, modifications, and variations will be apparent to those skilled in the art. In the claims, means-plus-function clauses are intended to cover the structures described herein as performing the recited function and not only structural equivalents but also equivalent structures.